The EMBO Journal vol.6 no.12 pp.3553–3558, 1987

# The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators

**Javier Paz-Ares, Debabrota Ghosal, Udo Wienand, Peter A.Peterson[1] and Heinz Saedler**

Max-Planck Institut für Züchtungsforschung, Egelspfad, D-5000 Köln 30, FRG and [1]Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Communicated by Heinz Saedler

**The structure of the wild-type *c1* locus of *Zea mays* was determined by sequence analysis of one genomic and two cDNA clones. The coding region is composed of three exons (150 bp, 129 bp and one, at least 720 bp) and two small introns (88 bp and 145 bp). Transcription of the mRNAs corresponding to the two cDNA clones cLC6 (1.1 kb) and cLC28 (2.1 kb) starts from the same promoter. Both cDNAs are identical except that cLC28 extends further at its 3′ end. A putative protein, 273 amino acids in length was deduced from the sequence of both transcripts. It contains two domains, one basic and the other acidic and might function as a transcriptional activator. The basic domain of this *c1*-encoded protein shows 40% sequence homology to the protein products of animal *myb* proto-oncogenes.**

*Key words: c1* locus/DNA-binding protein/*myb* proto-oncogenes/*Zea mays*

## Introduction

The biosynthesis of the purple plant pigment anthocyanin involves several enzymatic steps (for review see Coe and Neufer, 1977). In *Zea mays* (maize) a number of loci which affect anthocyanin biosynthesis have been identified by recessive mutations (Coe, 1957; Reddy and Coe, 1962; Dooner and Nelson, 1977, 1979). Some of these loci have been shown to encode enzymes involved in this pathway; these include *c2* (chalcone synthase: Dooner, 1983; Wienand *et al.*, 1986), *pr* (3′-hydroxylase: Larson, 1986), *a1* (NADPH-dependent reductase: Schwarz-Sommer *et al.*, 1987; Reddy *et al.*, 1987), and *bz1* (UDP-glucosyl-transferase: Larson and Coe, 1977; Dooner and Nelson, 1977). Other loci, such as *a2* and *bz2*, have not been allocated defined enzymatic functions, even though they are believed to encode for such activities based on precursor feeding and accumulation experiments (McCormick, 1978; Reddy and Coe, 1962).

The loci identified thus far, though not genetically linked, are expressed in a coordinated manner (Dooner, 1983). This is due to the activity of several genes regulating anthocyanin biosynthesis in various tissues of the plant. At least seven such regulatory loci have been identified genetically. Among these are *c1, r1 vp1, pl* and *clf*, each of which is essential for expression of at least the chalcone synthase and UDP-glucosyl-transferase activities (Dooner and Nelson, 1979; Dooner, 1983). Whereas the *c1* locus is required for pigmentation in the aleurone and scutellum of maize kernels, it does not affect anthocyanin production in other parts of the plant (Chen and Coe, 1977).

Recently the *c1* locus was cloned using a transposon tagging strategy (Paz-Ares *et al.*, 1986; Cone *et al.*, 1986) making it the first regulatory locus in plants to be available for detailed molecular analysis. Here we present the DNA sequence and the structure of the *c1* locus together with the sequence of two cDNAs which define a *c1*-encoded protein. This 29-kd protein has homology to the products of animal *myb* proto-oncogenes (Klempnauer *et al.*, 1982; Gonda *et al.*, 1985; Katzen *et al.*, 1985; Majello *et al.*, 1986). The protein encoded by the *c1* locus has a basic amino terminus and an acidic carboxy terminus, thereby potentially representing the two domains characteristic for transcriptional activators such as GAL4 and GCN4 gene products of yeast (for review see Struhl, 1987).

## Results

### *The DNA sequence of the c1 locus*

The *c1* locus was previously cloned from transposable element-induced mutants (Paz-Ares *et al.*, 1986; Cone *et al.*, 1986). The molecular analysis of seven mutant alleles showed that the locus extends over a region of at least 3 kb (Figure 1). The 5′-most insert is the *En1* element present in the mutant *c-m 668613* (insert 1 in Figure 1) and the 3′-most insert is the *Ds* element from the mutant *c1-m1* (insert 7 in Figure 1). *c1*-specific transcripts (Paz-Ares *et al.*, 1986; Cone *et al.*, 1986) map within the region bounded by inserts 1 and 7 (Figure 1), indicating that this portion of the wild-type clone most likely covers the entire *c1* locus.

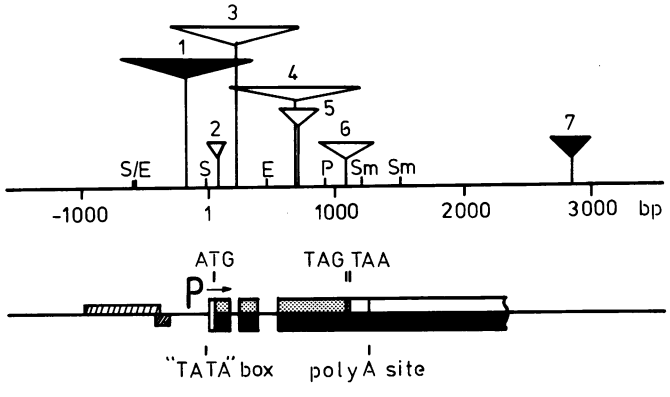


**Fig. 1.** Structure of the *c1* locus of maize. The upper part of the figure shows the molecularly defined positions of insertions in various *c1* mutants (**1, 4** and **5**: Paz-Ares *et al.*, 1986, **6**: Paz-Ares, unpublished; **2, 3, 5, 6** and **7**: Cone *et al.* 1986). These mutants are: 1 = *c-m 668613::En;* 2 = *c1-m858::dSpm*; 3 = *c1-m5::Spm*; 4 = *c-m668655::En*; 5 = *c1-m2::Ds*; 6= *C1-I*; and 7 = *c1-m1::Ds*. The outermost inserts (1 and 7) interfering with expression of the *c1* locus are highlighted in black. The lower part represents a schematic drawing of the structure of the locus according to the sequence data presented in Figure 2. The box upstream of the promoter (P) shows the region of homology between the *c1* and *c2* loci. The small box at its right end shows the position of the homology to the cDNA clones csLC1, csLC2 and csLC3. The three boxes downstream of the promoter represent the three exons defined by cDNAs cLC6 and cLC28. The upper part of the boxes (light shadow) show the translated part of the *c1* locus (as defined by the cDNA cLC6). The poly(A) site indicates the position of poly(A) addition. Exon three probably extends further 3′ in some transcripts but is shown as defined by the sequence of the longer truncated cDNA clone cLC28. Restriction sites are: E, *Eco*RI; S, *Sal*I; P, *Pst*I; SM, *Sma*I.

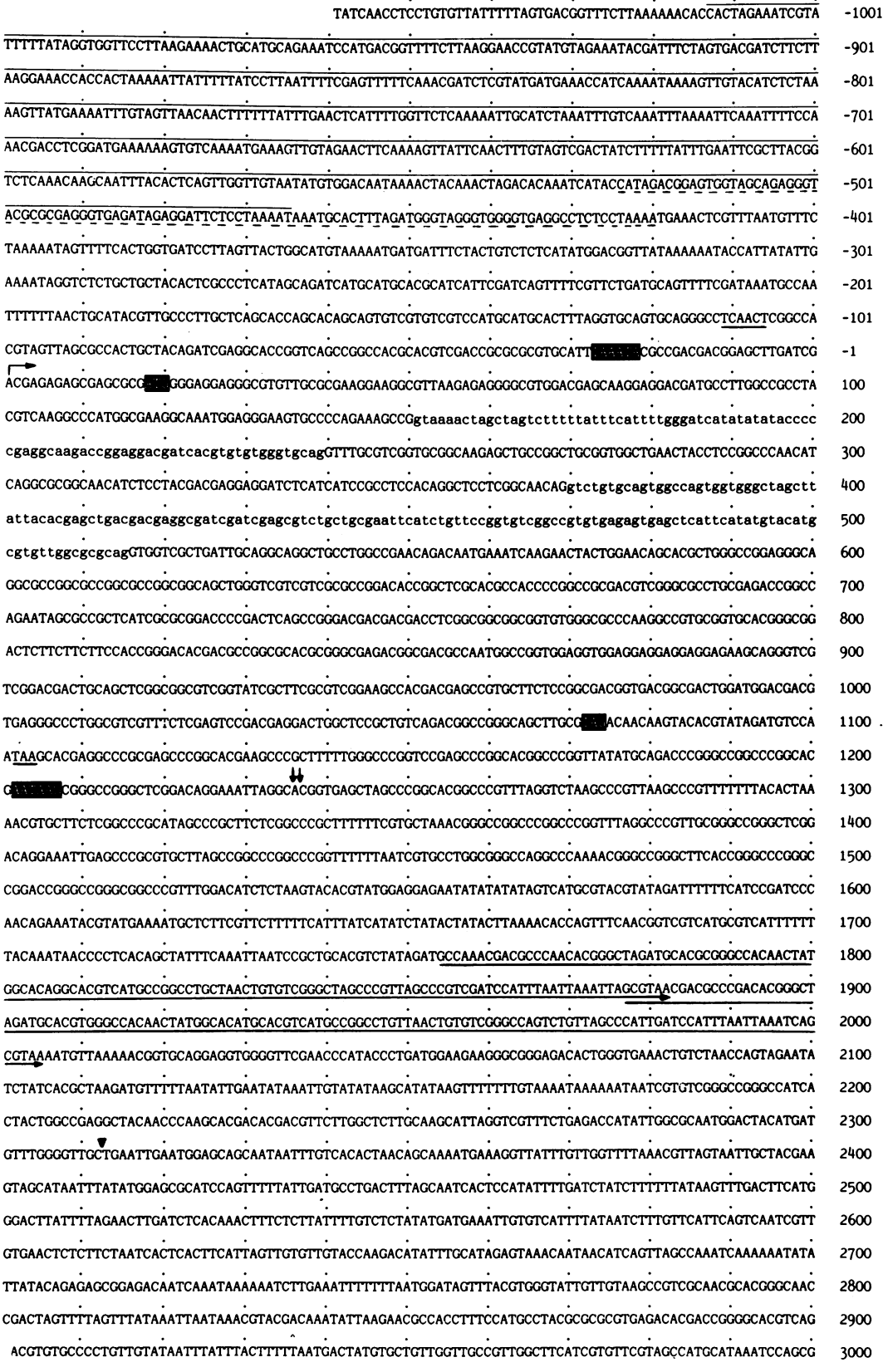

**Fig. 2.** Sequence of the *c1* locus of maize. Position 1 indicates the transcriptional start site as determined by S1 and primer extension experiments. The TATA box, the translational start, the stop codon and the poly(A) addition sequence are highlighted in black. The intron sequences according to cDNA clone cLC6 are given in small letters. The CAAT box at position −107 and the second stop codon TAA at position 1102 are underlined. Also underlined is the region of homology between the *c1* and *c2* loci (−1013 to −465). The dotted line at the c1 promoter proximal part of this homologous region represents sequences homologous to the small cDNAs csLC1, csLC2 and csLC3. Arrows at positions 1235 and 1236 indicate the poly(A) addition site of cDNA clone cLC6. The arrow at position 2312 shows where cDNA clone cLC28 is terminated. The long stretches of duplicated sequences within nucleotides 1754 and 2005 are underlined.

To determine the structure of the *c1* locus, a 5-kb fragment including the region between insert 1 and 7 (Figure 1) was subcloned and sequenced (Figure 2).

*The isolation of c1-specific cDNAs*

As described earlier (Paz-Ares *et al.*, 1986), three transcripts, 300 bp, 1.4 kb and 1.6 kb in size, hybridize with the 1-kb *Eco*RI fragment of *c1* (Figure 2, position −612 to +450). A further transcript, 2.5 kb in length, also hybridizes with the same probe (Cone *et al.*, 1986).

*Analysis of cDNAs homologous to the small transcript.* If the 570 bp long *Eco*RI−*Sal*I fragment (Figure 2 position −612 to position −42) is used as a probe in Northern blotting experiments with endosperm poly(A)$^+$ RNA, a 300-bp band is detected (Paz-Ares *et al.*, 1986). Three cDNA clones, csLC1, csLC2 and csLC3 homologous to this small RNA were isolated. Although each of the clones was unique, there was 80−86% DNA sequence homology among them (data not shown). Comparison of these sequences with the genomic sequence (Figure 2) demonstrated that none of the three cDNAs was originated from this genomic fragment. The homology of these cDNAs to the genomic sequence is confined to the area between positions −534 and −420 (indicated by the broken line in Figure 2). Furthermore, each clone is homologous to only a portion of this 115 bp long genomic sequence. The poly(A) end of the three small cDNAs is proximal to the 5′ end of the larger transcripts.

*Analysis of cDNA clones homologous to the large transcripts.* To clone cDNAs for large *c1* transcripts a cDNA library was established in λNM1149 from endosperm mRNA of the maize line C. This library was screened with the genomic 1-kb *Eco*RI fragment which contains the 5′ end of the gene (see Figure 1). Among the $10^6$ recombinant phages two homologous clones (cLC6 and cLC28) were obtained and sequenced (Figure 2). This low number of clones probably resulted from selection for clones containing the 5′ ends of the transcripts.

The 1.1-kb cDNA clone cLC6 is almost full-size, since it has a poly(A) tail (100 bp) and also contains 6 bp of the short 5′-untranslated sequence (Figure 2). At the 5′ end of this clones, an 81 bp long sequence representing the genomic sequence from position 724 to 804 is present in reverse orientation. As we believe, this is due to an artefact in cDNA cloning and is not considered as being part of the corresponding mRNA. cDNA clone cLC6 may represent the 1.4-kb *c1*-specific poly(A)$^+$ RNA seen previously in Northern blots (Paz-Ares *et al.*, 1986), because in those experiments the only size standard used was rRNA; the 1.4 kb size estimate may therefore have been an overestimation.

The second cDNA clone cLC28 is 2.1 kb in size, but does not appear to be full-size, because no poly(A) tail is present. The sequence of the 5′ part of cLC28 is identical to cLC6 except that cLC28 does not show the 81-bp repeat present in cLC6. Clone cLC28 extends beyond the 3′ end of cLC6 (Figures1 and 2). This cDNA could correspond to the 2.5-kb transcript described by Cone *et al.* (1986).

A cDNA clone corresponding to the *c1*-specific 1.6-kb transcript described earlier (Paz-Ares *et al.*, 1986) has not yet been isolated.

*The structure of the c1 locus*

A comparison of the genomic sequence of the *c1* locus with the sequences of the cDNA clones (cLC6 and cLC28) revealed the presence of two introns in the coding region of the gene (Figures 1 and 2). The first intron is 88 bp long and inserted between two G nucleotides of a Gly codon (amino acid 45) and the second intron is 144 bp long and insert between two G nucleotides

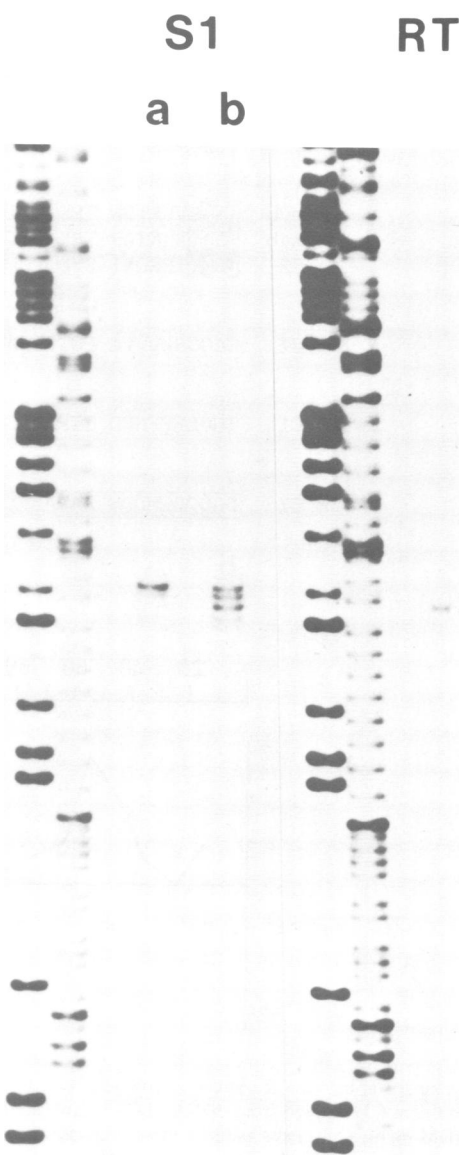


**Fig. 3.** Determination of the transcription start of *c1*-specific transcripts. A 727-bp *Eco*RI−*Nco*I fragment (position −612 to +115) was labeled at the 5′ end and used in the S1 experiment (S1) as well as for the sequencing reactions (G- and A-specific lanes are shown). A 30-bp synthetic oligonucleotide (position +85 to +115 in Figure 1) was 5′ labeled and then used in the primer extension experiment. The S1-protected fragments (S1) after 10 min (**a**) and 30 min (**b**) of treatment with the S1 enzyme and reverse transcriptase generated fragments (RT) were run together with the G- and A-specific reactions on an 8% sequencing gel. The sequence ladder corresponds to the strand complementary to the RNA.

of an Arg codon (amino acid 88). Both introns conform to the GT−AG rule of exon−intron borders (Breathnach and Chambon, 1981). The first exon is 150 bp and the second 129 bp in length. Since the two cDNAs cLC6 and cLC28 are identical over ~ 1 kb, exons 1 and 2 are the same for both transcripts. The length of exon 3 based on the sequence of cDNA clone cLC6 would be 720 bp while based on the sequence of cLC28 [which was truncated having no poly(A) tail] it would be at least 1796 bp.

The start of transcription was identified by S1 mapping and primer extension experiments. The data obtained in these experiments (Figure 3) indicated that *c1*-specific mRNAs transcribed from the region investigated all seemed to start with the same nucleotide (Figure 2, nucleotide A at position 1). cDNA clone

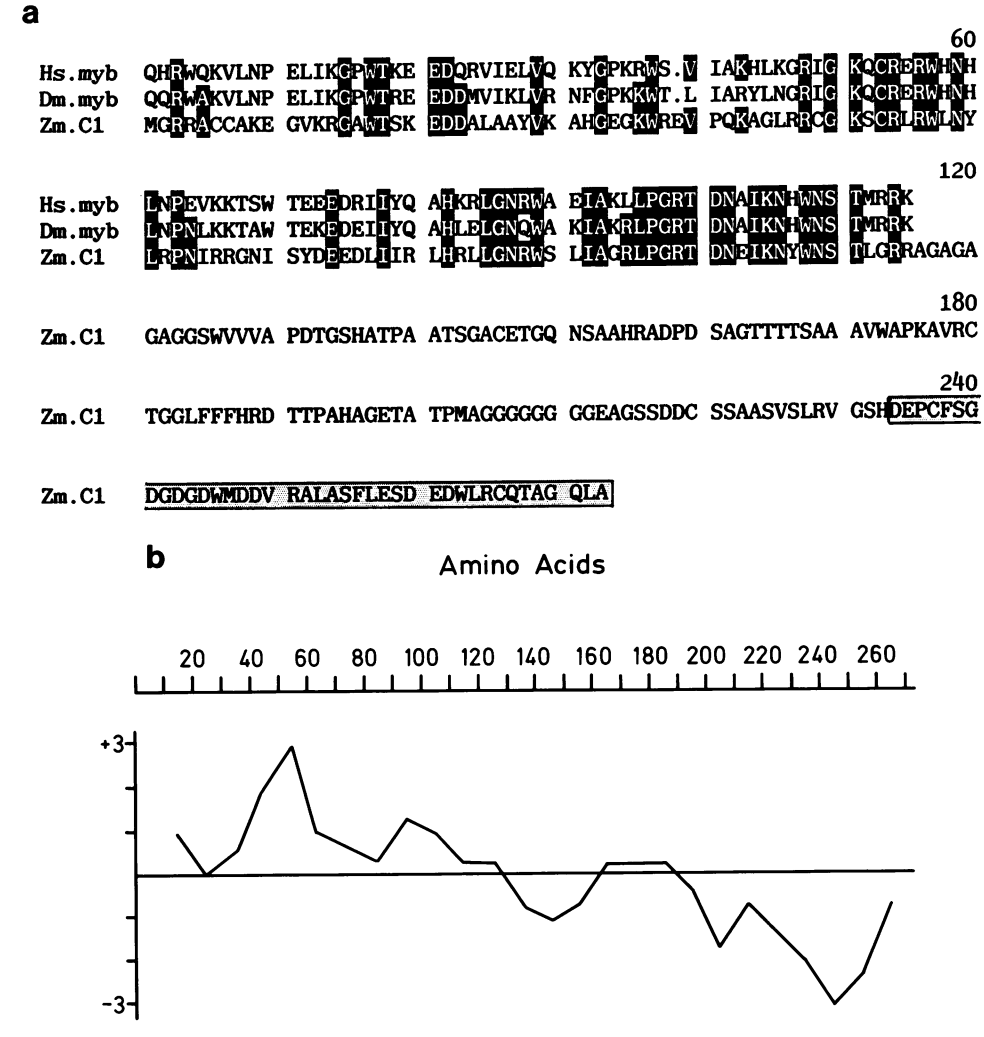


**Fig. 4.** (**a**) Comparison between the amino acid sequence of the putative *C1* protein and of the products of *myb* proto-oncogenes. The amino acid sequence of the *C1* protein (Zm.C1) derived from cDNA cLC6 is compared to the amino acid sequence of *myb* proto-oncogene products from humans (Hs.myb) and *Drosophila* (DM.myb). Only those parts of these proteins which show homology to the C1 protein (basic domain) are shown. Amino acids common to all three proteins are boxed in black. The carboxy terminus of the *C1* protein (from amino acid 234 to 273) which contains the acidic domain is boxed in. (**b**) Charge distribution in the putative *C1* protein. Shown is the average of net charges over successive 30 amino acids, measured at 10 amino acid intervals.

cLC6 starts with an A at position 12, while cLC28 begins with G at position 13. The leader sequence is rather short since the first ATG is located at position 18 and opens the longest open reading frame. This initiation codon is included in the sequence GCCATGG, in agreement with the Kozak rules for the initiation of translation (Kozak, 1984). A putative TATA box, TAAATA, is found upstream of the transcription start point at position −22 to −28. The sequence TCAACT, similar to the CAAT box (Breathnach and Chambon, 1981) precedes the TATA box at −111.

The first stop codon is a TAG at position 1072 (highlighted black in Figure 2), which is followed in-frame by a second stop codon TAA at position 1102 (underlined in Figure 2). The TAG codon is suppressed in the tobacco−TMV system (Beier *et al.*, 1984a) as well as in the wheat germ system (Beier *et al.*, 1984b). Whether TAG is also suppressed in maize and especially in the aleurone tissue is unknown. The possibility of suppression of TAG introduces some uncertainty about the exact termination of translation. For the smaller cDNA cLC6 there is a sequence, AATAAG, 126 bp downstream of the TAG codon (position 1202, Figure 2), resembling a putative polyadenylation signal (Proudfoot and Brownlee, 1976). This sequence is 27 (or 28) bp 5′ of the poly(A) addition site, depending on whether or not the A at position 1236 is considered part of the gene.

A 550-bp region extending from position −1013 to −465 in Figure 2 shows 80% homology to a region 1−1.5 kb upstream of the chalcone synthase structural gene (*c2*) of *Z. mays* (Wienand *et al.*, 1986), the expression of which is controlled by the *c1* locus (Dooner and Nelson, 1979; Dooner, 1983). Whether this region of homology is important in the regulation of the *c1* locus itself or in the *c1*-mediated control of anthocyanin biosynthetic genes, such as *c2*, *a1* or *bz1* is unknown. In Southern experiments we could not detect the presence of this sequence within 5 kb upstream of the promoter region of either the *a1* (Schwarz-Sommer *et al.*, 1987) or the *bz1* genes (D.Furtek, personal communication; Fedoroff *et al.*, 1984) (data not shown).

There are two ~400 bp long direct duplications of the *c1* locus sequence. One copy extends from position 1088 to at least 1499 and the second copy from position 3535 to at least 3924. These duplications are 90% homologous. The second copy of this sequence is downstream of the sequence shown in Figure 2.

A second set of direct duplications that are 90% homologous to each other starts at position 1754 and ends at position 1882, while the second copy overlaps the first by 5 bp at position 1877

and extends to position 2005. Any significance of these duplications remains unclear.

*Features of the c1-encoded protein*

The sequence of the putative protein encoded by the *c1* locus (Figure 4a) was derived by the theoretical translation of the cDNA sequences into the corresponding amino acid sequence. The protein has 273 amino acids and a mol. wt of ~29 kd. There are two domains, a basic one at the amino terminus between amino acids 1 and 114 and an acidic one at the carboxy terminus between amino acids 234 and 261 (Figure 4b). Comparison of this protein sequence with the protein sequences present in the NBRF data bank revealed a region between amino acids 2 and 114 homologous to the basic domain of *myb* proto-oncogenes. Figure 4a shows this region of homology to the protein products of *myb* proto-oncogenes from human and *Drosophila* (Majello *et al.*, 1986; Katzen *et al.*, 1985).

## Discussion

*The structure of the c1 locus*

The *c1* promoter region responsible for the transcription of the mRNAs homologous to cLC6 and cLC28 was defined at the 3′ end by the initiation of transcription at position +1 and putative TATA and CAAT boxes (Figures 3 and 2). A region spanning > 550 bp and showing 80% homology to a 5′ upstream region of the *c2* locus (a gene controlled by *c1*: Dooner and Nelson, 1979; Dooner, 1983) is located ~1 kb upstream from the TATA box. This region of homology seems not be be involved in the regulation of biosynthetic genes involved in pigmentation, since it could not be detected in the 5′ promoter region of the *a1* and *bz1* genes, which are also under control of *c1*. The region of homology between the *c1* and *c2* loci partially overlaps with the three cDNA sequences derived from the 300-bp transcripts. Whether the *c1* locus region homologous to the 300-bp transcripts (around position −500) is still part of the *c1* promoter is unknown. None of the cloned transposable element-induced *c1* mutants (Paz-Ares *et al.*, 1986; Cone *et al.*, 1986) have inserts in the region with homology to the small transcripts. The 300-bp mRNAs do not represent a high copy number sequence in the maize genome, since in Southern experiments only 15−20 copies homologous to these transcripts could be detected (data not shown).

The 3′ end of the *c1* locus is even less well defined. To data the most distal insert affecting *c1*-controlled phenotypes is the *Ds* element present in the mutant *c1-m1*, which, according to restriction mapping (Cone *et al.*, 1986), seems to be located around position 2800. This element is inserted ~1.7 kb downstream of the 3′ end of the cDNA cLC6. This *Ds* insert could affect the processing of the primary transcript corresponding to the 1.4-kb and 2.5-kb mRNAs, since the 3′ end of the latter could be proximal to this point of insertion. Processing of the 1.6-kb *c1* transcript could also be affected, if this mRNA is an alternatively spliced derivative of the same primary transcript. *Ds* elements in some instances have been shown to interfere with the splicing pattern (Simon and Starlinger, 1987). Alternatively the *Ds* insertion at *c-m1* could interfere with some regulatory sequence located in the 3′ end of the gene. The presence of regulatory sequences in the 3′ end of the gene has been reported for other genes like the adult β-globin and histone H5 genes from chicken (Choi and Engel, 1986; Trainor *et al.*, 1987) and the Adh2 gene of *Drosophila mulleri* (Fischer and Maniatis, 1986).

*The c1 locus might code for a transcriptional activator*

One function of the *c1* locus is suggested by the cDNA clone cLC6. The amino acid sequence of the putative protein derived from cLC6 has interesting features which indicate that the *c1* locus might encode a transcriptional activator. This 29-kb *C1* protein is 273 amino acids long. The amino-terminal part of the protein (Figure 4a; amino acids 1−114) shows a remarkable homology (40%) to the products of *myb* proto-oncogenes from human, chicken, mouse and *Drosophila* (Majello *et al.*, 1986; Gonda *et al.*, 1985; Katzen *et al.*, 1985; Figure 4a). These proto-oncogenes, whose functions are unknown, code for nuclear proteins (Klempnauer *et al.*, 1982; Boyle *et al.*, 1984) with DNA binding capacity (Moelling *et al.*, 1985; Klempnauer and Sipple, 1986, 1987; Klempnauer *et al.*, 1986). The DNA binding activity is located in the basic domain of these proteins (Klempnauer and Sippel, 1987). The homology of proteins encoded by animal *myb* proto-oncogenes and the protein encoded by the *c1* locus entends over this basic domain (Figure 4a) and might indicate that the *C1* protein is also a DNA-binding protein. Since the product of the *c1* locus seems to affect the expression of genes involved solely in anthocyanin biosynthesis it seems likely that the possible DNA binding activity of the protein encoded by the *c1* locus may be sequence specific. Experiments are in progress to establish whether the protein encoded by the *c1* locus is able to specifically bind sequences present in the genes under the control of *c1* such as *a1*, *c2* and *bz1*.

There is an acidic domain observed at the carboxy terminus of the protein encoded by the *c1* locus (amino acids 241−262, Figure 4a). Such short acidic domains have been shown to be important components of proteins involved in the activation of transcription in yeast. The protein products of these transcription activator genes (GCN4 and GAL4) contain small stretches of acidic amino acids at different positions with respect to the DNA-binding domains (Hope and Struhl, 1986; Ma and Ptashne, 1987). DNA binding and the activation of transcription in these proteins are two separable functions (Brent and Ptashne, 1985). The finding of a basic and an acidic domain of the *c1*-encoded protein could indicate that this protein is a DNA-binding protein which might activate transcription. This is likely because *c1* gene product(s) are required for *c2* and *bz1* gene expression (Dooner, 1983).

Whether the putative *c1* protein described here is the only protein product of the *c1* locus is not known. Genetic studies of the *Ds*-induced *c1* mutant *c-m2* (Figure 2, insert 5) suggest a bifunctional nature for *c1*, possibly involving the production of two substances, necessary for pigment formation (McClintock, 1949). If this is the case, one of the products could be the putative *c1*-encoded protein translated from the 1.4-kb mRNA and described here, while the second one might be derived from the larger 1.6-kb RNA. Further cDNA cloning followed by sequence analysis is needed to investigate this possibility.

*Evolutionary and functional implications*

The findings of homology between the product of the *c1* locus and the *myb* proteins demonstrates the presence of (proto)oncogene-related sequences in plants. This fact probably indicates that the ancestral *myb* gene was present before the divergence between animals and plants.

The observed homology of the proteins encoded by the *c1* locus and the *myb* proteins is in the same region as is the homology between the different vertebrate proteins. However, the degree of homology between maize and animals is only 40% compared with 75% homology between animals. This is in line with the

evolutionary relationships between the respective groups. It is unlikely that the products of the *c1* locus and *myb* proto-oncogenes serve similar physiological functions within their host organisms, since *c1* regulates a biochemical pathway not present in animals. In addition, mutations at the *c1* locus do not appear to affect maize development, except for the increase or decrease of pigmentation in the aleurone and scutellum tissues of the kernel. We consider that the observed homology between maize and animal *myb* oncogene products might reflect a similar mode of action for these proteins (i.e. regulation at the transcriptional level through DNA–protein interaction).

## Materials and methods

*Plant strains*

LC which was used as a source of the wild-type *c1* allele is a color-converted W22 maize line developed by Dr R.A.Brink, Wisconsin.

*Standard molecular procedures*

Plasmid and poly(A)$^+$ mRNA preparation was performed as previously described (Schwarz-Sommer *et al.*, 1984). cDNA cloning in λNM1149 followed the protocol of Schwarz-Sommer *et al.* (1985). A library of $10^6$ plaques derived from line C endosperm poly(A) RNA (30 days after pollination) was screened with the 1-kb *Eco*RI fragment described in the results section (Figures 1 and 2). Two of the positive clones cLC6 and cLC28 were analyzed and sequenced. Isolation of the small cDNA clones csLC1, csLC2 and csLC3 was done by constructing a cDNA library from endosperm poly(A)$^+$ RNA (see above) enriched for small mol. wt RNA by sucrose gradient centrifugation (Maniatis *et al.*, 1982). This library was then screened with the 1-kb *Eco*RI fragment of the wild-type *c1* clone number 5 (Paz-Ares *et al.*, 1986).

*S1 mapping*

A 727-bp *Eco*RI–*Nco*I fragment (corresponding to position −612 to +115 was 5′ end labeled at the *Nco*I site and used as a probe for S1 mapping. Ten μg of poly(A)$^+$ RNA isolated from developing maize kernels (from maize line C; 30 days after pollination) were precipitated together with 150 000 c.p.m. of the labeled fragment (30 ng). Annealing and S1 treatment were carried out following the method of Weaver and Weissmann (1979). The mixture was annealed at 54°C overnight and treated with S1 (300 units/ml) at 20°C for 10 min and 30 min. After precipitation with ethanol the pellet was treated with 0.4 M NaOH for 12 h at 25°C, neutralized with acetic acid and precipitated with ethanol. Analysis was then carried out on an 8% sequencing gel.

*Primer extension experiments*

A 5′-labeled single stranded 30-mer (position +85 to +115) was annealed to 10 μg of poly(A)$^+$ RNA as described in the S1 mapping. Excess of primer was removed by affinity chromatography on oligo(dT)–cellulose. After ethanol precipitation the mixture was dissolved and the primer extended for 1 h with 600 units of M-MLLV reverse transcriptase (BRL) following the instructions of the manufacturer. Analysis was then carried out on an 8% sequencing gel.

*Sequence analysis*

The DNA sequence of most of the genomic DNA and the 2.1-kb cDNA cLC28 was performed by the dideoxy chain termination method (Sanger *et al.*, 1977). After subcloning in the M13 mp18 and mp19 vectors (Norrander *et al.*, 1983). For deletion subcloning the method of Henikoff (1984) was followed. Minor parts of the genomic DNAs and the other cDNAs were sequenced by the chemical modification procedure (Maxam and Gilbert, 1980). All protein-encoding regions and 5′ promoter regions shown in Figure 1 were read from both strands. The 3′ non-coding region was read from at least two different subclones.

*Computer program for protein comparison*

The NBRF protein databank was screened for sequences homologous to the *c1* protein using the wordsearch program from the UW GCG program library (Devereux *et al.*, 1984).

## Acknowledgements

We thank Alfons Gierl, Derek Lydiate, Laurence Maréchal, William Martin, Pat Schnable, Brian Scheffler and Zsuzsanna Schwarz-Sommer for critical reading of the manuscript and helpful discussions. We also thank Ulla Niesback-Klösgen, Zsuzsanna Schwarz-Sommer and Douglas Furtek for providing clones and sequence data concerning the *c2*, *a1* and *bz1* gene. J.P.-A. was supported by a long-term EMBO post-doctoral fellowship.

## References

Beier,H., Barciszewska,M., Krupp,G., Mitnacht,R. and Gross,H.J. (1984a) *EMBO J.*, **3**, 351–356.
Beier,H., Barciszewska,M. and Sickinger,H.D. (1984b) *EMBO J.*, **3**, 1091–1096.
Boyle,W.J., Lambert,M.A., Lipsick,J.S. and Baluda,M.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4265–4269.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Brent,R. and Ptashne,M. (1985) *Cell*, **43**, 729–736.
Chen,S.M. and Coe,E.H.,Jr (1977) *Biochem. Genet.*, **15**, 333–346.
Choi,O.R. and Engel,J.D. (1986) *Nature*, **323**, 731–734.
Coe,E.H.,Jr (1957) *Am. Nat.*, **91**, 381–385.
Coe,E.H.,Jr and Neufer,G.M. (1977) In Sprague,G.F. (ed.), *Corn and Corn Improvement*. The American Society of Agronomy Inc., Madison, WI, pp. 111–223.
Cone,K.C., Burr,F.A. and Burr,B. (1986) *Proc Natl. Acad. Sci. USA*, **83**, 9631–9635.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Dooner,H.K. (1983) *Mol. Gen. Genet.*, **189**, 136–141.
Dooner,H.K. and Nelson,O.E. (1977) *Biochem. Genet.*, **15**, 509–515.
Dooner,H.K. and Nelson,O.E. (1979) *Genetics*, **91**, 309–315.
Fedoroff,N.V., Furtek,D.B. and Nelson,E.N.,Jr (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3825–3829.
Fischer,J.A. and Maniatis,T. (1986) *EMBO J.*, **5**, 1275–1289.
Gonda,T.J., Gough,N.M., Dunn,A.R. and Blaquiere,J. (1985) *EMBO J.*, **4**, 2003–2008.
Henikoff,S. (1984) *Gene*, **28**, 351–359.
Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.
Katzen,A.L., Kornberg,T.B. and Bishop,J.M. (1985) *Cell*, **41**, 449–456.
Klempnauer,K.H. and Sippel,A.E. (1986) *Mol. Cell. Biol.*, **6**, 62–69.
Klempnauer,K.H. and Sippel, A.E. (1987) *EMBO J.*, **6**, 2719–2725.
Klempnauer.K., Gonda,T.J. and Bishop,J.M. (1982) *Cell*, **31**, 453–463.
Klempnauer,K.H., Bonifer,C. and Sippel,A.E. (1986) *EMBO J.*, **5**, 1903–1911.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Larson,R. (1986) *Maize Gen. Coop. News Lett.*, **60**, 48–49.
Larson,R. and Coe,E.H.,Jr (1977) *Biochem. Genet.*, **15**, 153–156.
Ma,J. and Ptashne,M. (1987) *Cell*, **48**, 847–853.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Majello,B., Kenyon,L.C. and Dalla-Favera,R.D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9636–9640.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Moelling,K., Pfaff,E., Beng,P.B., Bunte,T., Schaller,H.E. and Graf,T. (1985) *Cell*, **40**, 983–990.
McClintock,B. (1949) *Yearbook '48*. Carnegie Institute of Washington. pp. 142–154.
McCormick,S. (1978) *Biochem. Genet.*, **16**, 777–785.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Paz-Ares,J., Wienand,U., Peterson,P.A. and Saedler,H. (1986) *EMBO J.*, **5**, 829–833.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Reddy,G.M. and Coe,E.H.,Jr (1962) *Science*, **138**, 149–150.
Reddy,A., Salamini,F., Saedler,H. and Rohde,W. (1987) *Plant Sci.*, **52**, in press.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5464.
Schwarz-Sommer,Zs., Gierl,A., Klösgen,R.B., Wienand,U., Peterson,P.A. and Saedler,H. (1984) *EMBO J.*, **3**, 1021–1028.
Schwarz-Sommer,Zs., Gierl,A., Cuypers,H., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 591–597.
Schwarz-Sommer,Zs., Shepherd,N., Tacke,E., Gierl,A., Rohde,W., Lequlerc,L., Mattes,M., Berntgen,R., Peterson,P.A. and Saedler,H. (1987) *EMBO J.*, **6**, 287–294.
Simon,R. and Starlinger,P. (1987) *Maize Coop. News Lett.*, **61**, 42.
Struhl,K. (1987) *Cell*, **49**, 295–297.
Trainor,C.D., Stamler,S.J. and Engel,J.D. (1987) *Nature*, **328**, 827–830.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175–1193.
Wienand,U., Weydemann,U., Niesbach-Klösgen,U., Peterson,P.A. and Saedler,H. (1986) *Mol. Gen. Genet.*, **203**, 202–207.

*Received on July 16, 1987; revised on September 11, 1987*